Application Data Sheet

Application Information

Application number::	09/930,020
Filing Date::	08/14/01
Application Type::	Regular
Subject Matter::	Utility
Suggested classification::	
Suggested Group Art Unit::	
CD-ROM or CD-R??::	
Number of CD disks::	
Number of copies of CDs::	
Sequence Submission::	
Computer Readable Form (CRF)?::	
Number of copies of CRF::	
Title::	Methods of Diagnosis of Colorectal Cancer, Compositions and Methods of Screening for Colorectal Cancer Modulators
Attorney Docket Number::	018501-003100US
Request for Early Publication::	No
Request for Non-Publication::	No
Suggested Drawing Figure::	
Total Drawing Sheets::	0
Small Entity?::	Yes
Latin name::	
Variety denomination name::	
Petition included?::	No
Petition Type::	
Licensed US Govt. Agency::	
Contract or Grant Numbers One::	
Secrecy Order in Parent Appl.::	No

Initial 10/16/01

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Kurt
Middle Name::	C.
Family Name::	Gish
Name Suffix::	
City of Residence::	San Francisco
State or Province of Residence::	CA
Country of Residence::	US
Street of Mailing Address::	40 Perego Terrace, No. 2
City of Mailing Address::	San Francisco
State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94131

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	David
Middle Name::	H.
Family Name::	Mack
Name Suffix::	
City of Residence::	Menlo Park
State or Province of Residence::	CA
Country of Residence::	US
Street of Mailing Address::	2076 Monterey Avenue
City of Mailing Address::	Menlo Park
State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94025

Initial 10/16/01

Applicant Authority Type::	Inventor
Primary Citizenship Country::	US
Status::	Full Capacity
Given Name::	Keith
Middle Name::	E.
Family Name::	Wilson
Name Suffix::	
City of Residence::	Redwood City
State or Province of Residence::	CA
Country of Residence::	US
Street of Mailing Address::	219 Jeter Street
City of Mailing Address::	Redwood City
State or Province of mailing address::	CA
Country of mailing address::	US
Postal or Zip Code of mailing address::	94062

Correspondence Information

Correspondence Customer Number::   20350

Representative Information

Representative Customer Number::   20350

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::
09/633,733	CIP	Utility	September 15, 2000

Foreign Priority Information

Country::	Application number::	Filing Date::

Assignee Information

Assignee Name::   Eos Biotechnology, Inc.
Street of mailing address::   225A Gateway Boulevard
City of mailing address::   South San Francisco
State or Province of mailing address::   California
Country of mailing address::   USA
Postal or Zip Code of mailing address::   94080

Initial 10/16/01

Preliminary Amendments

3

Level - 2
Version 1.1
Updated - 8/01/01

On November 9, 2001

TOWNSEND and TOWNSEND and CREW LLP

By: _Jill R. Clarke_

Jill R. Clarke

COPY OF PAPERS
ORIGINALLY FILED

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

GISH *et al.*

Application No.: 09/930,020

Filed: August 14, 2001

For: METHODS OF DIAGNOSIS OF COLORECTAL CANCER, COMPOSITIONS AND METHODS OF SCREENING FOR COLORECTAL CANCER MODULATORS

Examiner: Not yet assigned

Art Unit: 1642

COMMUNICATION UNDER

37 C.F.R. §§ 1.821-1.825

AND

PRELIMINARY AMENDMENT

Box SEQUENCE
Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In response to the request to comply with Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures, 37 C.F.R. §§ 1.821-1.825, that accompanied the Notice to File Missing Parts of Nonprovisional Application mailed September 13, 2001, Applicants submit herewith the required paper copy and computer readable copy of the Sequence Listing. Please amend the specification as follows.

In the Specification:

Please replace paragraph **[45]** beginning at page 13, line 9 with the following:

--**[45]** The extracellular domains of transmembrane proteins are diverse; however, conserved motifs are found repeatedly among various extracellular domains. Conserved structure and/or functions have been ascribed to different extracellular motifs. For example, cytokine receptors are characterized by a cluster of cysteines and a WSXWS (SEQ ID NO:3) (W= tryptophan, S= serine, X=any amino acid) motif. Immunoglobulin-like domains are highly conserved. Mucin-like domains may be involved in cell adhesion and leucine-rich repeats participate in protein-protein interactions.--

Please replace the paragraph (**TABLE 2**) beginning at page 96, line 5 with the following:

--TABLE 2 CBF9 DNA and Protein Sequences

CBF9 DNA sequence (SEQ ID NO:1)

Gene name: ESTs
Unigene number: Hs.157601
Probeset Accession #: W07459
Nucleic Acid Accession #: AC005383
Coding Sequence: 328-2751 (underlined sequences correspond to start and stop codons)

```
1         11         21         31         41         51
|          |          |          |          |          |
GACAGTGTTC GCGGCTGCAC CGCTCGGAGG CTGGGTGACC CGCGTAGAAG TGAAGTACTT    60
TTTTATTTGC AGACCTGGGC CGATGCCGCT TTAAAAAACG CGAGGGGCTC TATGCACCTC   120
CCTGGCGGTA GTTCCTCCGA CCTCAGCCGG GTCGGGTCGT GCCGCCCTCT CCCAGGAGAG   180
ACAAACAGGT GTCCCACGTG GCAGCCGCGC CCCGGGCGCC CCTCCTGTGA TCCCGTAGCG   240
CCCCCTGGCC CGAGCCGCGC CCGGGTCTGT GAGTAGAGCC GCCCGGGCAC CGAGCGCTGG   300
TCGCCGCTCT CCTTCCGTTA TATCAAC ATG CCCCCTTTCC TGTTGCTGGA GGCCGTCTGT   360
GTTTTCCTGT TTTCCAGAGT GCCCCCATCT CTCCCTCTCC AGGAAGTCCA TGTAAGCAAA   420
GAAACCATCG GGAAGATTTC AGCTGCCAGC AAAATGATGT GGTGCTCGGC TGCAGTGGAC   480
ATCATGTTTC TGTTAGATGG GTCTAACAGC GTCGGGAAAG GGAGCTTTGA AAGGTCCAAG   540
CACTTTGCCA TCACAGTCTG TGACGGTCTG GACATCAGCC CCGAGAGGGT CAGAGTGGGA   600
GCATTCCAGT TCAGTTCCAC TCCTCATCTG GAATTCCCCT TGGATTCATT TTCAACCCAA   660
CAGGAAGTGA AGGCAAGAAT CAAGAGGATG GTTTTCAAAG GAGGGCGCAC GGAGACGGAA   720
CTTGCTCTGA AATACCTTCT GCACAGAGGG TTGCCTGGAG GCAGAAATGC TTCTGTGCCC   780
CAGATCCTCA TCATCGTCAC TGATGGGAAG TCCCAGGGGG ATGTGGCACT GCCATCCAAG   840
CAGCTGAAGG AAAGGGGTGT CACTGTGTTT GCTGTGGGGG TCAGGTTTCC CAGGTGGGAG   900
GAGCTGCATG CACTGGCCAG CGAGCCTAGA GGGCAGCACG TGCTGTTGGC TGAGCAGGTG   960
GAGGATGCCA CCAACGGCCT CTTCAGCACC CTCAGCAGCT CGGCCATCTG CTCCAGCGCC  1020
ACGCCAGACT GCAGGGTCGA GGCTCACCCC TGTGAGCACA GGACGCTGGA GATGGTCCGG  1080
GAGTTCGCTG GCAATGCCCC ATGCTGGAGA GGATCGCGGC GGACCCTTGC GGTGCTGGCT  1140
GCACACTGTC CCTTCTACAG CTGGAAGAGA GTGTTCCTAA CCCACCCTGC CACCTGCTAC  1200
AGGACCACCT GCCCAGGCCC CTGTGACTCG CAGCCCTGCC AGAATGGAGG CACATGTGTT  1260
CCAGAAGGAC TGGACGGCTA CCAGTGCCTC TGCCCGCTGG CCTTTGGAGG GGAGGCTAAC  1320
TGTGCCCTGA AGCTGAGCCT GGAATGCAGG GTCGACCTCC TCTTCCTGCT GGACAGCTCT  1380
GCGGGCACCA CTCTGGACGG CTTCCTGCGG GCCAAAGTCT TCGTGAAGCG GTTTGTGCGG  1440
GCCGTGCTGA GCGAGGACTC TCGGGCCCGA GTGGGTGTGG CCACATACAG CAGGGAGCTG  1500
CTGGTGGCGG TGCCTGTGGG GGAGTACCAG GATGTGCCTG ACCTGGTCTG GAGCCTCGAT  1560
GGCATTCCCT TCCGTGGTGG CCCCACCCTG ACGGGCAGTG CCTTGCGGCA GGCGGCAGAG  1620
CGTGGCTTCG GGAGCGCCAC CAGGACAGGC CAGGACCGGC CACGTAGAGT GGTGGTTTTG  1680
CTCACTGAGT CACACTGGGA GGATGAGGTT GCGGGCCCAG CGCGTCACGC AAGGGCGCGA  1740
GAGCTGCTCC TGCTGGGTGT AGGCAGTGAG GCCGTGCGGG CAGAGCTGGA GGAGATCACA  1800
GGCAGCCCAA AGCATGTGAT GGTCTACTCG GATCCTCAGG ATCTGTTCAA CCAAATCCCT  1860
GAGCTGCAGG GGAAGCTGTG CAGCCGGCAG CGGCCAGGGT GCCGGACACA AGCCCTGGAC  1920
CTCGTCTTCA TGTTGGACAC CTCTGCCTCA GTAGGGCCCG AGAATTTGC TCAGATGCAG  1980
AGCTTTGTGA GAAGCTGTGC CCTCCAGTTT GAGGTGAACC CTGACGTGAC ACAGGTCGGC  2040
CTGGTGGTGT ATGGCAGCCA GGTGCAGACT GCCTTCGGGC TGGACACCAA ACCCACCCGG  2100
GCTGCGATGC TGCGGGCCAT TAGCCAGGCC CCCTACCTAG GTGGGGTGGG CTCAGCCGGC  2160
ACCGCCCTGC TGCACATCTA TGACAAAGTG ATGACCGTCC AGAGGGGTGC CCGGCCTGGT  2220
```

```
GTCCCCAAAG CTGTGGTGGT GCTCACAGGC GGGAGAGGCG CAGAGGATGC AGCCGTTCCT   2280
GCCCAGAAGC TGAGGAACAA TGGCATCTCT GTCTTGGTCG TGGGCGTGGG GCCTGTCCTA   2340
AGTGAGGGTC TGCGGAGGCT TGCAGGTCCC CGGGATTCCC TGATCCACGT GGCAGCTTAC   2400
GCCGACCTGC GGTACCACCA GGACGTGCTC ATTGAGTGGC TGTGTGGAGA AGCCAAGCAG   2460
CCAGTCAACC TCTGCAAACC CAGCCCGTGC ATGAATGAGG GCAGCTGCGT CCTGCAGAAT   2520
GGGAGCTACC GCTGCAAGTG TCGGGATGGC TGGGAGGGCC CCCACTGCGA GAACCGTGAG   2580
TGGAGCTCTT GCTCTGTATG TGTGAGCCAG GGATGGATTC TTGAGACGCC CCTGAGGCAC   2640
ATGGCTCCCG TGCAGGAGGG CAGCAGCCGT ACCCCTCCCA GCAACTACAG AGAAGGCCTG   2700
GGCACTGAAA TGGTGCCTAC CTTCTGGAAT GTCTGTGCCC CAGGTCCTTA_GAATGTCTGC   2760
TTCCCGCCGT GGCCAGGACC ACTATTCTCA CTGAGGGAGG AGGATGTCCC AACTGCAGCC   2820
ATGCTGCTTA GAGACAAGAA AGCAGCTGAT GTCACCCACA AACGATGTTG TTGAAAAGTT   2880
TTGATGTGTA AGTAAATACC CACTTTCTGT ACCTGCTGTG CCTTGTTGAG GCTATGTCAT   2940
CTGCCACCTT TCCCTTGAGG ATAAACAAGG GGTCCTGAAG ACTTAAATTT AGCGGCCTGA   3000
CGTTCCTTTG CACACAATCA ATGCTCGCCA GAATGTTGTT GACACAGTAA TGCCCAGCAG   3060
AGGCCTTTAC TAGAGCATCC TTTGGACGGC GAAGGCCACG GCCTTTCAAG ATGGAAAGCA   3120
GCAGCTTTTC CACTTCCCCA GAGACATTCT GGATGCATTT GCATTGAGTC TGAAAGGGGG   3180
CTTGAGGGAC GTTTGTGACT TCTTGGCGAC TGCCTTTTGT GTGTGGAAGA GACTTGGAAA   3240
GGTCTCAGAC TGAATGTGAC CAATTAACCA GCTTGGTTGA TGATGGGGGA GGGGCTGAGT   3300
TGTGCATGGG CCCAGGTCTG GAGGGCCACG TAAAATCGTT CTGAGTCGTG AGCAGTGTCC   3360
ACCTTGAAGG TCTTC
```

CBF9 Protein sequence (SEQ ID NO:2)

Gene name:	ESTs
Unigene number:	Hs.157601

Protein Accession #:   none found

Signal sequence:	1-17
Transmembrane domains:	none found
VGW domains:	49-223; 341-518; 529-706
EGF domains:	298-333; 715-748
Cellular Localization:	plasma membrane

```
1          11         21         31         41         51
|          |          |          |          |          |
MPPFLLLEAV CVFLFSRVPP SLPLQEVHVS KETIGKISAA SKMMWCSAAV DIMFLLDGSN   60
SVGKGSFERS KHFAITVCDG LDISPERVRV GAFQFSSTPH LEFPLDSFST QQEVKARIKR  120
MVFKGGRTET ELALKYLLHR GLPGGRNASV PQILIIVTDG KSQGDVALPS KQLKERGVTV  180
FAVGVRFPRW EELHALASEP RGQHVLLAEQ VEDATNGLFS TLSSSAICSS ATPDCRVEAH  240
PCEHRTLEMV REFAGNAPCW RGSRRTLAVL AAHCPFYSWK RVFLTHPATC YRTTCPGPCD  300
SQPCQNGGTC VPEGLDGYQC LCPLAFGGEA NCALKLSLEC RVDLLFLLDS SAGTTLDGFL  360
RAKVFVKRFV RAVLSEDSRA RVGVATYSRE LLVAVPVGEY QDVPDLVWSL DGIPFRGGPT  420
LTGSALRQAA ERGFGSATRT GQDRPRRVVV LLTESHSEDE VAGPARHARA RELLLLGVGS  480
EAVRAELEEI TGSPKHVMVY SDPQDLFNQI PELQGKLCSR QRPGCRTQAL DLVFMLDTSA  540
SVGPENFAQM QSFVRSCALQ FEVNPDVTQV GLVVYGSQVQ TAFGLDTKPT RAAMLRAISQ  600
APYLGGVGSA GTALLHIYDK VMTVQRGARP GVPKAVVVLT GGRGAEDAAV PAQKLRNNGI  660
SVLVVGVGPV LSEGLRRLAG PRDSLIHVAA YADLRYHQDV LIEWLCGEAK QPVNLCKPSP  720
CMNEGSCVLQ NGSYRCKCRD GWEGPHCENR EWSSCSVCVS QGWILETPLR HMAPVQEGSS  780
RTPPSNYREG LGTEMVPTFW NVCAPGP
```

Please insert the accompanying paper copy of the Sequence Listing, page numbers 1 to 4, at the end of the application.
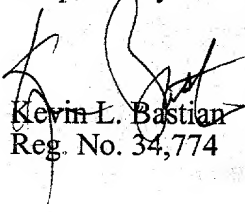
REMARKS

Applicants request entry of this amendment in adherence with 37 C.F.R. §§1.821 to 1.825. This amendment is accompanied by a floppy disk containing the above named sequences, SEQ ID NOS:1-3, in computer readable form, and a paper copy of the sequence information which has been printed from the floppy disk.

The information contained in the computer readable disk was prepared through the use of the software program "PatentIn" and is identical to that of the paper copy. This amendment contains no new matter.

Attached hereto is a marked-up version of the changes made to the Specification and Abstract by the current Amendment. The attached pages are captioned **"VERSION WITH MARKINGS TO SHOW CHANGES MADE."**

If the Examiner believes a telephone conference would expedite prosecution of this application, please telephone the undersigned at 415-576-0200.

Respectfully submitted,

Kevin L. Bastian
Reg. No. 34,774

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: (415) 576-0200
Fax: (415) 576-0300
KLB:dmw

VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the Specification:

Paragraph **[45]** beginning at line 23 of page 6 has been amended as follows:

     [45]   The extracellular domains of transmembrane proteins are diverse; however, conserved motifs are found repeatedly among various extracellular domains. Conserved structure and/or functions have been ascribed to different extracellular motifs. For example, cytokine receptors are characterized by a cluster of cysteines and a WSXWS (SEQ ID NO:3) (W= tryptophan, S= serine, X=any amino acid) motif. Immunoglobulin-like domains are highly conserved. Mucin-like domains may be involved in cell adhesion and leucine-rich repeats participate in protein-protein interactions.

Paragraph (**TABLE 2**) beginning at line 5 of page 96 has been amended as follows:

TABLE 2 CBF9 DNA and Protein Sequences

<u>CBF9 DNA sequence</u> (SEQ ID NO:1)

Gene name: ESTs
Unigene number: Hs.157601
Probeset Accession #:   W07459
Nucleic Acid Accession #:   AC005383
Coding Sequence: 328-2751 (underlined sequences correspond to start
and stop codons)


```
 1          11          21          31          41          51
 |          |           |           |           |           |
GACAGTGTTC GCGGCTGCAC CGCTCGGAGG CTGGGTGACC CGCGTAGAAG TGAAGTACTT    60
TTTTATTTGC AGACCTGGGC CGATGCCGCT TTAAAAAACG CGAGGGGCTC TATGCACCTC   120
CCTGGCGGTA GTTCCTCCGA CCTCAGCCGG GTCGGGTCGT GCCGCCCTCT CCCAGGAGAG   180
ACAAACAGGT GTCCCACGTG GCAGCCGCGC CCCGGGCGCC CCTCCTGTGA TCCCGTAGCG   240
CCCCCTGGCC CGAGCCGCGC CCGGGTCTGT GAGTAGAGCC GCCCGGGCAC CGAGCGCTGG   300
TCGCCGCTCT CCTTCCGTTA TATCAAC<u>ATG</u> CCCCCTTTCC TGTTGCTGGA GGCCGTCTGT   360
GTTTTCCTGT TTTCCAGAGT GCCCCCATCT CTCCCTCTCC AGGAAGTCCA TGTAAGCAAA   420
GAAACCATCG GGAAGATTTC AGCTGCCAGC AAAATGATGT GGTGCTCGGC TGCAGTGGAC   480
ATCATGTTTC TGTTAGATGG GTCTAACAGC GTCGGGAAAG GGAGCTTTGA AAGGTCCAAG   540
CACTTTGCCA TCACAGTCTG TGACGGTCTG GACATCAGCC CCGAGAGGGT CAGAGTGGGA   600
GCATTCCAGT TCAGTTCCAC TCCTCATCTG GAATTCCCCT TGGATTCATT TTCAACCCAA   660
CAGGAAGTGA AGGCAAGAAT CAAGAGGATG GTTTTCAAAG GAGGGCGCAC GGAGACGGAA   720
CTTGCTCTGA AATACCTTCT GCACAGAGGG TTGCCTGGAG GCAGAAATGC TTCTGTGCCC   780
CAGATCCTCA TCATCGTCAC TGATGGGAAG TCCCAGGGGG ATGTGGCACT GCCATCCAAG   840
CAGCTGAAGG AAAGGGGTGT CACTGTGTTT GCTGTGGGGG TCAGGTTTCC CAGGTGGGAG   900
GAGCTGCATG CACTGGCCAG CGAGCCTAGA GGGCAGCACG TGCTGTTGGC TGAGCAGGTG   960
GAGGATGCCA CCAACGGCCT CTTCAGCACC CTCAGCAGCT CGGCCATCTG CTCCAGCGCC  1020
ACGCCAGACT GCAGGGTCGA GGCTCACCCC TGTGAGCACA GGACGCTGGA GATGGTCCGG  1080
GAGTTCGCTG GCAATGCCCC ATGCTGGAGA GGATCGCGGC GGACCCTTGC GGTGCTGGCT  1140
GCACACTGTC CCTTCTACAG CTGGAAGAGA GTGTTCCTAA CCCACCCTGC CACCTGCTAC  1200
AGGACCACCT GCCCAGGCCC CTGTGACTCG CAGCCCTGCC AGAATGGAGG CACATGTGTT  1260
CCAGAAGGAC TGGACGGCTA CCAGTGCCTC TGCCCGCTGG CCTTTGGAGG GGAGGCTAAC  1320
TGTGCCCTGA AGCTGAGCCT GGAATGCAGG GTCGACCTCC TCTTCCTGCT GGACAGCTCT  1380
GCGGGCACCA CTCTGGACGG CTTCCTGCGG GCCAAAGTCT TCGTGAAGCG GTTTGTGCGG  1440
GCCGTGCTGA GCGAGGACTC TCGGGCCCGA GTGGGTGTGG CCACATACAG CAGGGAGCTG  1500
CTGGTGGCGG TGCCTGTGGG GGAGTACCAG GATGTGCCTG ACCTGGTCTG GAGCCTCGAT  1560
GGCATTCCCT TCCGTGGTGG CCCCACCCTG ACGGGCAGTG CCTTGCGGCA GGCGGCAGAG  1620
CGTGGCTTCG GGAGCGCCAC CAGGACAGGC CAGGACCGGC CACGTAGAGT GGTGGTTTTG  1680
CTCACTGAGT CACACTCCGA GGATGAGGTT GCGGGCCCAG CGCGTCACGC AAGGGCGCGA  1740
GAGCTGCTCC TGCTGGGTGT AGGCAGTGAG GCCGTGCGGG CAGAGCTGGA GGAGATCACA  1800
GGCAGCCCAA AGCATGTGAT GGTCTACTCG GATCCTCAGG ATCTGTTCAA CCAAATCCCT  1860
GAGCTGCAGG GGAAGCTGTG CAGCCGGCAG CGGCCAGGGT GCCGGACACA AGCCCTGGAC  1920
CTCGTCTTCA TGTTGGACAC CTCTGCCTCA GTAGGGCCCG AGAATTTTGC TCAGATGCAG  1980
AGCTTTGTGA GAAGCTGTGC CCTCCAGTTT GAGGTGAACC CTGACGTGAC ACAGGTCGGC  2040
CTGGTGGTGT ATGGCAGCCA GGTGCAGACT GCCTTCGGGC TGGACACCAA ACCCACCGG   2100
GCTGCGATGC TGCGGGCCAT TAGCCAGGCC CCCTACCTAG GTGGGGTGGG CTCAGCCGGC  2160
ACCGCCCTGC TGCACATCTA TGACAAAGTG ATGACCGTCC AGAGGGGTGC CCGGCCTGGT  2220
```

```
GTCCCCAAAG CTGTGGTGGT GCTCACAGGC GGGAGAGGCG CAGAGGATGC AGCCGTTCCT 2280
GCCCAGAAGC TGAGGAACAA TGGCATCTCT GTCTTGGTCG TGGGCGTGGG GCCTGTCCTA 2340
AGTGAGGGTC TGCGGAGGCT TGCAGGTCCC CGGGATTCCC TGATCCACGT GGCAGCTTAC 2400
GCCGACCTGC GGTACCACCA GGACGTGCTC ATTGAGTGGC TGTGTGGAGA AGCCAAGCAG 2460
CCAGTCAACC TCTGCAAACC CAGCCCGTGC ATGAATGAGG GCAGCTGCGT CCTGCAGAAT 2520
GGGAGCTACC GCTGCAAGTG TCGGGATGGC TGGGAGGGCC CCCACTGCGA GAACCGTGAG 2580
TGGAGCTCTT GCTCTGTATG TGTGAGCCAG GGATGGATTC TTGAGACGCC CCTGAGGCAC 2640
ATGGCTCCCG TGCAGGAGGG CAGCAGCCGT ACCCCTCCCA GCAACTACAG AGAAGGCCTG 2700
GGCACTGAAA TGGTGCCTAC CTTCTGGAAT GTCTGTGCCC CAGGTCCTTA GAATGTCTGC 2760
TTCCCGCCGT GGCCAGGACC ACTATTCTCA CTGAGGGAGG AGGATGTCCC AACTGCAGCC 2820
ATGCTGCTTA GAGACAAGAA AGCAGCTGAT GTCACCCACA AACGATGTTG TTGAAAAGTT 2880
TTGATGTGTA AGTAAATACC CACTTTCTGT ACCTGCTGTG CCTTGTTGAG GCTATGTCAT 2940
CTGCCACCTT TCCCTTGAGG ATAAACAAGG GGTCCTGAAG ACTTAAATTT AGCGGCCTGA 3000
CGTTCCTTTG CACACAATCA ATGCTCGCCA GAATGTTGTT GACACAGTAA TGCCCAGCAG 3060
AGGCCTTTAC TAGAGCATCC TTTGGACGGC GAAGGCCACG GCCTTTCAAG ATGGAAAGCA 3120
GCAGCTTTTC CACTTCCCCA GAGACATTCT GGATGCATTT GCATTGAGTC TGAAAGGGGG 3180
CTTGAGGGAC GTTTGTGACT TCTTGGCGAC TGCCTTTTGT GTGTGGAAGA GACTTGGAAA 3240
GGTCTCAGAC TGAATGTGAC CAATTAACCA GCTTGGTTGA TGATGGGGGA GGGGCTGAGT 3300
TGTGCATGGG CCCAGGTCTG GAGGGCCACG TAAAATCGTT CTGAGTCGTG AGCAGTGTCC 3360
ACCTTGAAGG TCTTC
```

CBF9 Protein sequence (SEQ ID NO:2)

Gene name: ESTs
Unigene number: Hs.157601

Protein Accession #:   none found

Signal sequence: 1-17
Transmembrane domains: none found
VGW domains: 49-223; 341-518; 529-706
EGF domains: 298-333; 715-748
Cellular Localization: plasma membrane

```
  1          11          21          31          41          51
  |          |           |           |           |           |
MPPFLLLEAV CVFLFSRVPP SLPLQEVHVS KETIGKISAA SKMMWCSAAV DIMFLLDGSN   60
SVGKGSFERS KHFAITVCDG LDISPERVRV GAFQFSSTPH LEFPLDSFST QQEVKARIKR  120
MVFKGGRTET ELALKYLLHR GLPGGRNASV PQILIIVTDG KSQGDVALPS KQLKERGVTV  180
FAVGVRFPRW EELHALASEP RGQHVLLAEQ VEDATNGLFS TLSSSAICSS ATPDCRVEAH  240
PCEHRTLEMV REFAGNAPCW RGSRRTLAVL AAHCPFYSWK RVFLTHPATC YRTTCPGPCD  300
SQPCQNGGTC VPEGLDGYQC LCPLAFGGEA NCALKLSLEC RVDLLFLLDS SAGTTLDGFL  360
RAKVFVKRFV RAVLSEDSRA RVGVATYSRE LLVAVPVGEY QDVPDLVWSL DGIPFRGGPT  420
LTGSALRQAA ERGFGSATRT GQDRPRRVVV LLTESHSEDE VAGPARHARA RELLLLGVGS  480
EAVRAELEEI TGSPKHVMVY SDPQDLFNQI PELQGKLCSR QRPGCRTQAL DLVFMLDTSA  540
SVGPENFAQM QSFVRSCALQ FEVNPDVTQV GLVVYGSQVQ TAFGLDTKPT RAAMLRAISQ  600
APYLGGVGSA GTALLHIYDK VMTVQRGARP GVPKAVVVLT GGRGAEDAAV PAQKLRNNGI  660
SVLVVGVGPV LSEGLRRLAG PRDSLIHVAA YADLRYHQDV LIEWLCGEAK QPVNLCKPSP  720
CMNEGSCVLQ NGSYRCKCRD GWEGPHCENR EWSSCSVCVS QGWILETPLR HMAPVQEGSS  780
RTPPSNYREG LGTEMVPTFW NVCAPGP
```